

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re Patent Application of )  
Rainer KLISCH et al. ) Group Art Unit: 2641  
Application No.: 09/964,381 ) Examiner: Unassigned  
Filed: September 28, 2001 )  
For: METHOD AND DEVICE FOR )  
ANALYZING A SPOKEN SEQUENCE )  
OF NUMBERS )

**CLAIM FOR CONVENTION PRIORITY**

Assistant Commissioner for Patents  
Washington, D.C. 20231

Sir:

The benefit of the filing date of the following prior foreign application in the following foreign country is hereby requested, and the right of priority provided in 35 U.S.C. § 119 is hereby claimed:

European Patent Application No. 00121468.3

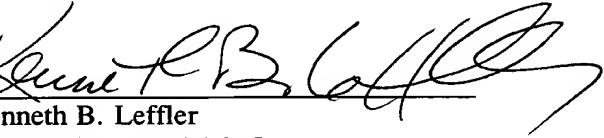
Filed: September 29, 2000

In support of this claim, enclosed is a certified copy of said prior foreign application. Said prior foreign application was referred to in the oath or declaration. Acknowledgment of receipt of the certified copy is requested.

Respectfully submitted,

BURNS, DOANE, SWECKER & MATHIS, L.L.P.

Date: November 27, 2001

By:   
Kenneth B. Leffler  
Registration No. 36,075

P.O. Box 1404  
Alexandria, Virginia 22313-1404  
(703) 836-6620

**THIS PAGE BLANK (USPTO)**

#5



**Europäisches  
Patentamt**

**European  
Patent Office**

**Office européen  
des brevets**



**Bescheinigung**

**Certificate**

**Attestation**

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

**Patentanmeldung Nr.      Patent application No.      Demande de brevet n°**

00121468.3

Der Präsident des Europäischen Patentamts;  
Im Auftrag

For the President of the European Patent Office  
Le Président de l'Office européen des brevets  
p.o.

I.L.C. HATTEN-HECKMAN

DEN HAAG, DEN  
THE HAGUE,  
LA HAYE, LE  
04/09/01



**THIS PAGE BLANK (USPTO)**



Europäisches  
Patentamt

European  
Patent Office

Office européen  
des brevets

**Blatt 2 der Bescheinigung  
Sheet 2 of the certificate  
Page 2 de l'attestation**

Anmeldung Nr.:  
Application no.: **00121468.3**  
Demande n°:

Anmeldetag:  
Date of filing: **29/09/00**  
Date de dépôt:

Anmelder:  
Applicant(s):  
Demandeur(s):  
**TELEFONAKTIEBOLAGET LM ERICSSON (publ)**  
**126 25 Stockholm**  
**SWEDEN**

Bezeichnung der Erfindung:  
Title of the invention:  
Titre de l'invention:  
**Method and device for analyzing a spoken sequence of numbers**

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat: State: Pays:	Tag: Date: Date:	Aktenzeichen: File no. Numéro de dépôt:
---------------------------	------------------------	---

Internationale Patentklassifikation:  
International Patent classification:  
Classification internationale des brevets:  
**G10L15/04, G10L15/26**

Am Anmeldetag benannte Vertragstaaten:  
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE/TR  
Etats contractants désignés lors du dépôt:

Bemerkungen:  
Remarks:  
Remarques:

**THIS PAGE BLANK (USPTO)**

EPO - Munich  
48  
29. Sep. 2000**Method and Device for Analyzing a Spoken Sequence of Numbers**

5

**BACKGROUND OF THE INVENTION****Technical Field**

10 The invention relates to a method and a device for analyzing a spoken sequence of numbers.

**Discussion of the Prior Art**

15 A lot of technical applications require recognition of a spoken sequence of numbers. Many mobile telephones comprise the feature of voice dialing by uttering a telephone number. Moreover, electronic commerce applications require the recognition of spoken order numbers and spoken credit card numbers.

20 WO-A-89 04035 discloses a method for recognizing a number like a telephone number consisting of a plurality of digits. The digits are uttered singly or in sequences. Two utterances comprising one or more digits may be separated by the user-defined placement of pauses. A pause time between two utterances is 25 monitored and when an utterance is followed by a pre-determined pause time interval, the recognized digits will be replied via a speech synthesizer. A further utterance comprising one or more digits can then be started, and only the next utterance will be replied after a subsequent pause.

30

While recognition of spoken digits and spoken digit sequences works reliably also under adverse noise conditions, automatic recognition of naturally spoken numbers like "twenty two" or "five hundred thirty" is more difficult. This is due to the 35 fact that spoken sequences of numbers like "twenty two" or "five hundred thirty" can stand for more than one numerical value. The spoken sequence of numbers "twenty two", for example, can stand either for the single numerical value "22" or for the two numerical values "20" and "2". As another example,

the sequence "five hundred thirty" can stand both for the numerical value "530" or for the two numerical values "500" and "30".

5 When automatically recognizing a spoken sequence of numbers, the recognition process becomes increasingly difficult if numbers with a large numerical value or a large sequence of numbers have to be analyzed. Thus, the spoken sequence of numbers "thousand four hundred fifty six" can stand for a single numerical value or for up to five numerical values. Altogether, 10 there exist eight possibilities: "1456", "1000" and "4" and "100" and "50" and "6", "1000" and "456", "1000" and "400" and "56", "1000" and "400" and "50" and "6", "1400" and "56", "1400" and "50" and "6", "1450" and "6".

15 These ambiguities do not only occur in the English language. In the German language , for example, the naturally spoken sequence of numbers "einhundert zehn" can stand both for the single numerical value "110" and the two numerical values "100" and "10". However, the ambiguities relating to the one or more numerical values of a spoken sequence of numbers may be different in different languages. While e. g. in the French language "quarante sept" can stand for both the single numerical value "47" or the two numerical values "40" and "7", this ambiguity 20 does not occur in the German language. In the German language the numerical value "47" is spoken as "siebenundvierzig" and the sequence of the two numerical values "40" and "7" is spoken as "vierzig sieben".

25 There is, therefore, a need for a method and device for analyzing a spoken sequence of numbers which allow a robust distinction between different semantic interpretations with respect to the one or more numerical values comprised therein.

## SUMMARY OF THE INVENTION

The present invention satisfies this need by providing a method  
5 for analyzing a spoken sequence of numbers, wherein the numbers  
are recognized by automatic speech recognition and wherein the  
method comprises determining a pause length between two con-  
secutive numbers and deciding whether or not the two consecu-  
tive numbers belong to a single numerical value on the basis of  
10 the determined pause length. A device for analyzing a spoken  
sequence of numbers comprises an automatic speech recognizer, a  
prosodic unit for determining a pause length between two con-  
secutive numbers and a processing unit for deciding whether or  
not the two consecutive numbers belong to a single numerical  
15 value on the basis of the determined pause length.

According to the invention, the speaking pause length between  
two consecutively spoken numbers is used as the single prosodic  
20 criterion or as one of a plurality of prosodic criteria for as-  
sessing whether or not the two consecutively spoken numbers be-  
long to a single numerical value or to two different numerical  
values. The speaking pause length is a robust prosodic crite-  
rion for analyzing a spoken sequence of numbers. Further pro-  
sodic parameters apart from the speaking pause length on which  
25 the decision whether or not two consecutively spoken numbers  
belong to a single numerical value can be based are known from  
E. Nöth et al "Prosodische Information: Begriffsbestimmung und  
Nutzen für das Sprachverständen", in Paulus, Wahl (ed.), Muster-  
erkennung 1997, Informatik aktuell, Springer-Verlag, Heidel-  
30 berg, 1997, pages 37-52, herewith incorporated by reference.

The decision whether or not two consecutively spoken numbers  
belong to a single numerical value can be a "hard" decision or  
a "soft" decision. The "hard" decision can be based on deter-  
35 mining whether or not certain thresholds of prosodic parameters  
have been exceeded. A "soft" decision may be made by means of a  
so-called classifier, e.g. a neuronal network, which takes into

account a plurality of prosodic parameters and which produces e.g. a probability decision.

According to a preferred embodiment of the invention, it is  
5 automatically decided that two consecutive numbers do not belong to a single numerical value if a certain pause length threshold is exceeded. Such a mechanism corresponds to the acoustical perception of a human listener. The two spoken numbers "20" and "2" e. g. will clearly be perceived by the human  
10 listener as two separate numerical values (i. e. "20" and "2") if a speaking pause of sufficient duration is made between speaking the numbers "20" and "2". On the other hand, the spoken numbers "20" and "2" will be perceived as a single numerical value (i. e. "22") if no or almost no speaking pause is  
15 made.

The speaking pause length threshold which forms the basis for the decision whether or not two consecutive numbers belong to a single numerical value can initially be set to a certain value.  
20 This value can be an empirical value estimated on the basis of a representative speech database. The pause length threshold can also be adjustable. This allows a user to adapt the speaking pause length threshold to his own manner-of-speaking, e. g. by changing the threshold value in system settings of the device.  
25

It has been found that robust setting of a pause length threshold is strongly interrelated with speech tempo which in turn depends on the individual speaker. In reality, the speech tempo of different speakers can vary within a wide range. According  
30 to a preferred embodiment of the invention, the pause length threshold is therefore automatically adapted to the current user's speaking habit. This can e. g. be done by analyzing previously determined speaking pause lengths within one or more  
35 previously uttered numerical values which the user has already acknowledged to be correct. A new pause length threshold can then either be set to the mean or the median computed over these previously determined speaking pause lengths or it can be

set anywhere between the old threshold and the mean or median value of the previously determined speaking pause lengths. In other words: the pause length threshold is shifted.

5       The decision whether or not two consecutively spoken numbers belong to a single numerical value can be made more robust if the decision is not only based on the speaking pause length but also on the previously mentioned further prosodic parameters apart from the speaking pause length. These further prosodic

10      parameters can relate to a phoneme duration like phrase-final lengthening or pre-boundary lengthening, the shape of the energy contour or specific pitch movements like phrase-final fall. Preferably, respective thresholds are also provided for these further prosodic parameters. The decision whether or not

15      two consecutive numbers belong to a single numerical value can accordingly also be based on the criterion whether or not a respective threshold of a further prosodic parameter has been exceeded.

20      Like the pause length threshold, the respective thresholds of further prosodic parameters can be user-adjustable or be automatically adjusted dependent on the user's speaking habit or be adjusted in accordance with appropriate training data. Moreover, previously determined further prosodic parameters of previously uttered numerical values which the user has already acknowledged to be correct can be used for shifting respective thresholds of the prosodic parameters.

30      In many languages, connecting words between two consecutive numbers of a spoken sequence of numbers indicate that the two consecutive numbers belong to one numerical value. In the English language, e. g., such a connecting word is the word "and". Thus, the spoken sequence of numbers "one hundred and ten" usually stands for the numerical value "110", even if the total pause length between "hundred" and "ten", the pause length between "hundred" and "and" or the pause length between "and" and "ten" exceeds a previously set pause length threshold.

In order to correctly analyze a spoken sequence of numbers comprising one or more connecting words between two consecutive numbers, a preferred embodiment of the invention comprises the feature of recognizing such a connecting word. According to a 5 first variant of the invention, it is determined that two consecutive numbers belong to a single numerical value every time a connecting word is arranged between the two numbers.

According to a second variant, upon recognition of a connecting 10 word between two consecutive numbers, the pause length threshold for determining whether or not the two consecutive numbers belong to a single numerical value is changed. In other words: upon recognition of a connecting word, the decision whether or not two consecutive numbers belong to a single nu- 15 mercial value is based on a different pause length threshold as in case no such connecting word is recognized. Consequently, two different pause length thresholds are utilized. Analyzing a spoken sequence of numbers thus becomes more robust because in certain cases the consecutive numbers belong to different nu- 20 mercial values although a connecting word is arranged therebe-tween, especially in cases where the pause length between the two consecutive numbers is extremely long (e. g. when a user places long pauses between the connecting word and the number preceding or following the connecting word).

25 There exist several possibilities for determining a speaking pause length between two consecutive numbers of a spoken se- 30 quence of numbers. The pause length can e. g. be directly de- termined by measuring a silence interval between two consecutively spoken numbers. This can be done with a so-called voice activity detector. A speaking pause length can also be determined indirectly using the information obtained as a by- 35 product from the process of automatic speech recognition. Dur- ing automatic speech recognition not only the words themselves but also their respective start and end points on a time axis are computed. The pause length can thus be determined based on an end point of the first of two consecutive numbers and a starting point of a second of two consecutive numbers. Espe-

cially in noisy environments, this technique usually leads to more robust results than measuring a silence interval between two consecutive numbers.

5

#### BRIEF DESCRIPTION OF THE DRAWINGS

Further aspects and advantages of the invention will become apparent upon reading the following detailed description of preferred embodiments of the invention and upon reference to the 10 drawings in which:

Fig. 1 is a schematic diagram of a device for analyzing a spoken sequence of numbers according to the invention; and

15

Fig. 2 is a schematic diagram of a method for analyzing a spoken sequence of numbers according to the invention.

20

#### DESCRIPTION OF THE PREFERERED EMBODIMENTS

In Fig. 1, a schematic diagram of a device 100 for analyzing a spoken sequence of numbers according to the invention is illustrated. The analyzing device 100 depicted in Fig. 1 comprises 25 an automatic speech recognizer 120, a prosodic unit 140 for determining a pause length between two consecutive numbers, a processing unit 160 for deciding if the two consecutive numbers belong to a single numerical value and an input unit 180.

Upon speaking a sequence of numbers like "five hundred thirty", the automatic speech recognizer 120 recognizes each of the spoken numbers as well as connecting words comprised within the spoken sequence of numbers. During the recognition process, the starting and end points in time of the recognized numbers and 35 connecting words are computed. These starting and end points are output to the prosodic unit 140 for determining the pause length between two consecutive numbers or between a connecting word and a preceding or subsequent number.

The processing unit 160 receives input from both the automatic speech recognizer 120 and the prosodic unit 140. Based on the numbers recognized by the automatic speech recognizer 120, the presence of connecting words between two consecutive numbers and the pause length between two consecutive numbers or a connecting word and a number preceding or following the connecting word, the processing unit 160 analyzes the spoken sequence of numbers with respect to the one or more numerical values contained therein.

The processing unit 160 decides whether or not two consecutive numbers belong to a single numerical value on the basis of a pause length threshold. This pause length threshold is initially set to a value between 100 ms and 1 s, preferably to a value of 200 ms.

By means of an input unit 180 a user has the possibility to adapt this initial threshold to his own manner-of-speaking. The input unit 180 comprises a graphical or physical slide bar allowing to adjust the threshold within a predetermined range. The input unit 180 also allows selection of an automatic adaptation of the threshold to the speaking habit of one or more users of the device 100.

The function of the device 100 is hereinafter described in more detail with reference to Fig. 2.

First of all, a pause length threshold  $\Theta$  is set automatically or by the user or according to appropriate training data to a certain value. Then, the user speaks the sequence "five hundred thirty" consisting of the three numbers "five", "hundred" and "thirty". These spoken numbers are subjected to automatic speech recognition in the automatic recognizer 120. The automatic speech recognizer 120 recognizes the three numbers "five", "hundred" and "thirty" with their respective starting and end points. The detection of the respective starting and end points indicates that there is a first pause between the

first number "five" and the second number "hundred" and a second pause between the second number "hundred" and the third number "thirty".

5 The starting and end points of the three numbers are input to the prosodic unit 140 which determines a pause length P1 of the first pause as well as a pause length P2 of the second pause. The three numbers recognized by the automatic speech recognizer 120 and the two pause lengths P1 and P2 determined by the prosodic unit 140 are input to the processing unit 160 which decides if two consecutive numbers belong to a single numerical value on the basis of the measured pause lengths P1 and P2.

15 If both the pause length P1 and the pause length P2 exceed the pause length threshold  $\Theta$ , the processing unit 160 decides that the spoken sequence of numbers contains three numerical values, i. e. "5", "100" and "30". If neither of the two pause lengths P1 and P2 exceeds the pause length threshold  $\Theta$ , the processing unit 160 decides that the spoken sequence of numbers contains a 20 single numerical value, i. e. "530".

If the processing unit 160 determines that only the first pause length P1 exceeds the pause length threshold  $\Theta$ , it decides that the spoken sequence of numbers contains the two numerical values "5" and "130". On the other hand, if only the second pause length P2 exceeds the pause length threshold  $\Theta$ , the processing unit 160 decides that the spoken sequence of numbers contains the two numerical values "500" and "30".

30 According to the method depicted in Fig. 2, the pause length P1 is determined prior to the pause length P2. This allows to analyze the spoken sequence of numbers in the order the numbers are spoken. Of course, the pause lengths P1 and P2 may also be determined and analyzed in a different order. This may necessitate that all numbers of the sequence of numbers have to be 35 spoken prior to the analyzing step.

Although the method depicted in Fig. 2 relates to a decision which is solely based on the determined pause length, the prosodic unit 140 depicted in Fig. 1 may also determine further prosodic parameters apart from the pause length and the decision may also be based on these further prosodic parameters.

5        Besides, the automatic speech recognizer 120 may also recognize connecting words within a spoken sequence of numbers and the processing unit 160 may, upon recognition of a connecting word, apply a different threshold regarding the one or more prosodic parameters on which the decision is based. Also, the decision can be based solely on one or more prosodic parameters apart 10      from the pause length.

The device 100 and the method according to the invention may be 15      used for many applications, e. g. stationary electronic commerce systems or mobile applications like mobile telephones.

EPO - Munich  
48  
29. Sep. 2000

## CLAIMS

- 5     1. A method for analyzing a spoken sequence of numbers recognized by automatic speech recognition, comprising:
  - determining a speaking pause length between two consecutive numbers; and
  - deciding whether or not the two consecutive numbers belong to a single numerical value on the basis of the determined pause length.
- 10    2. The method according to claim 1, further comprising determining one or more further prosodic parameters apart from the pause length and deciding whether or not the two consecutive numbers belong to a single numerical value based also on the one or more further prosodic parameters.
- 15    3. The method according to claim 1 or 2, wherein the decision is based on a threshold of the pause length and/or of the one or more further prosodic parameters.
- 20    4. The method according to claim 3, wherein the threshold is initially set to an empirical value.
- 25    5. The method according to claim 3 or 4, wherein the threshold is user-adjustable.
- 30    6. The method according to claim 3 or 4, wherein the threshold is automatically adjusted dependent on a user's speaking habit or dependent on appropriate training data.
- 35    7. The method according to one of claims 2 to 6, wherein the threshold of the pause length and/or of the further prosodic parameters is shifted on the basis of one or more previously determined pause lengths and/or previously determined further prosodic parameters relating to one or more correctly determined numerical values.

8. The method according to one of claims 1 to 7, wherein the pause length is determined by measuring a silence interval between two consecutive numbers.

5

9. The method according to one of claims 1 to 7, further comprising obtaining an end point of a first of the two consecutive numbers and a starting point of a second of the two consecutive numbers during automatic speech recognition and determining the pause length based on the end point and the starting point.

10

15

10. The method according to one of claims 1 to 9, further comprising recognizing a connecting word within the spoken sequence of numbers.

20

11. The method according to claim 10, wherein, upon recognition of a connecting word, the decision whether or not two consecutive numbers belong to a single numerical value is based on a different pause length threshold.

12. A device (100) for analyzing a spoken sequence of numbers comprising:

25

- an automatic speech recognizer (120);
- a prosodic unit (140) for determining a speaking pause length between two consecutive numbers; and
- a processing unit (160) for deciding whether or not the two consecutive numbers belong to a single numerical value on the basis of the determined pause length.

30

35

13. The device according to claim 12, wherein the prosodic unit (140) determines one or more further prosodic parameters apart from the speaking pause length and wherein the processing unit (160) decides whether or not the two consecutive numbers belong to a single numerical value based also on the one or more further prosodic parameters.

14. The device according to claim 12 or 13, wherein the automatic speech recognizer (120) recognizes a connecting word between the spoken sequence of numbers.

5

6069

**THIS PAGE BLANK (USPTO)**

Telefonaktiebolaget LM Ericsson (publ)  
P13726

- 14 -

EP-85 112

ABSTRACT

EPO - Munich  
48  
29. Sep. 2000

Method and Device for Analyzing a Spoken Sequence of Numbers

5 A method for analyzing a spoken sequence of numbers recognized by automatic speech recognition comprises determining the speaking pause length between two consecutive numbers and deciding if the two consecutive numbers belong to a single numerical value on the basis of the determined pause length. A  
10 device for analyzing a spoken sequence of numbers comprises an automatic speech recognizer, a unit for determining the pause length between two consecutive numbers and a processing unit for deciding if the two consecutive numbers belong to a single numerical value on the basis of the determined pause length.

15

(Fig. 2)

**THIS PAGE BLANK (USPTO)**

EPO - Munich  
48  
29. Sep. 2000

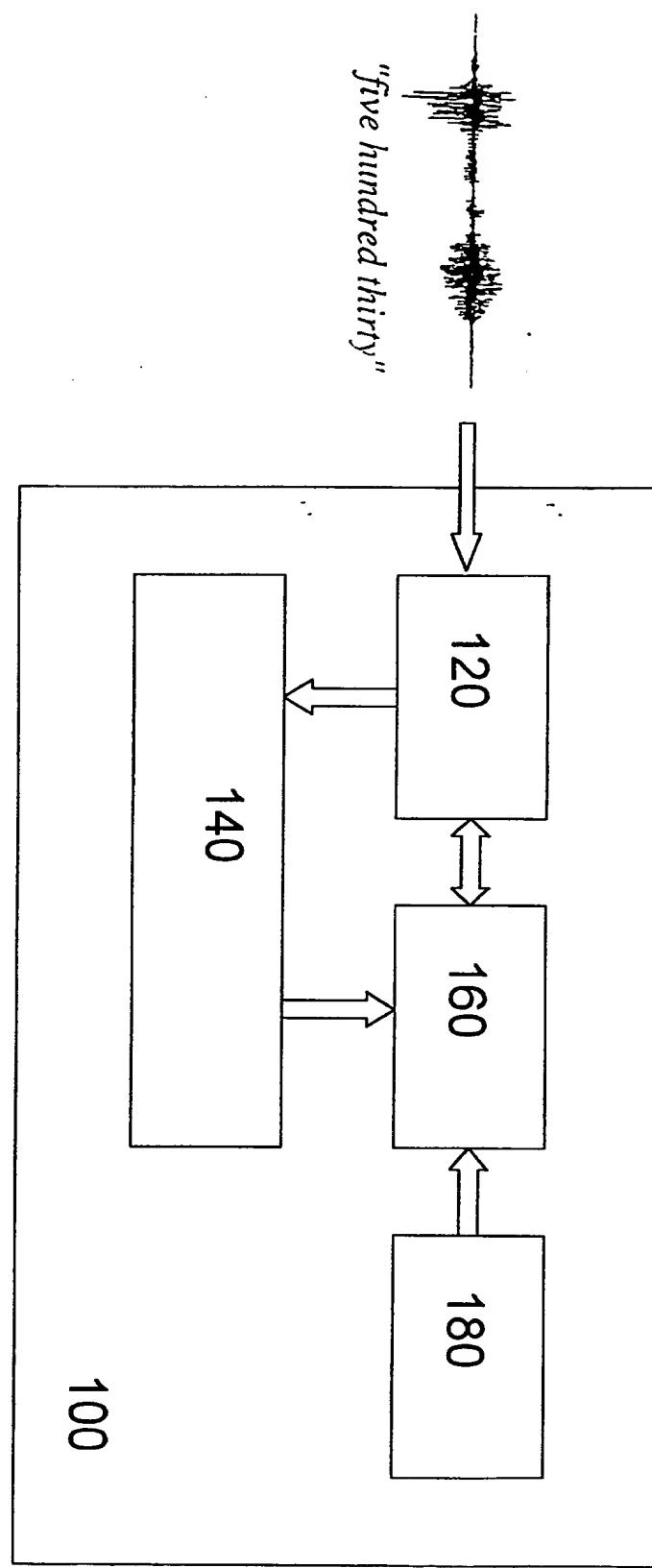


Figure 1

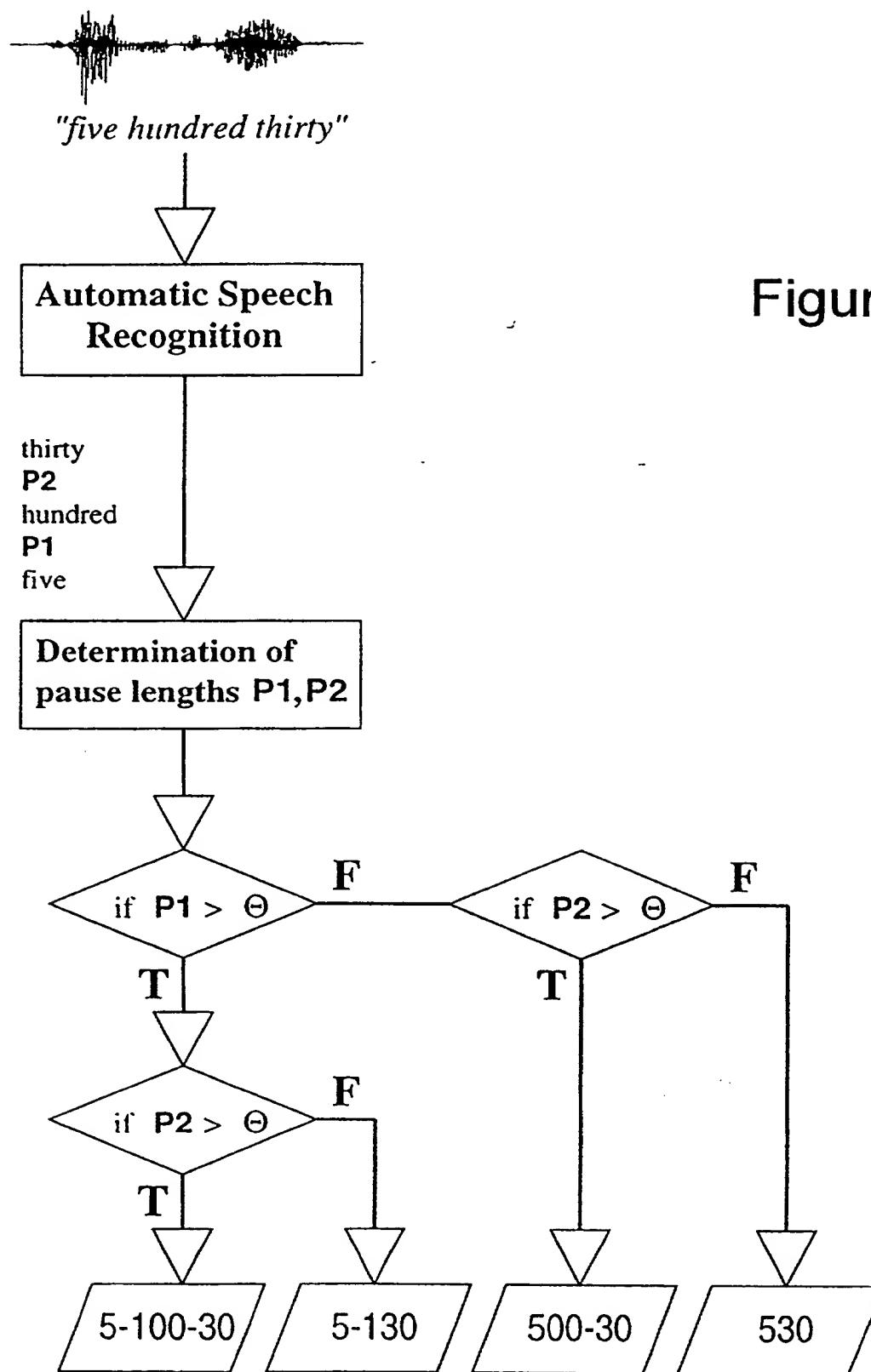


Figure 2